International Journal of Forecasting 16 (2000) 173–190
international journal of forecasting
www.elsevier.com/locate/ijforecast

Forecasting stock indices: a comparison of classification and level estimation models

Mark T. Leung[a,*], Hazem Daouk[b], An-Sing Chen[c]

[a]Department of Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, IN 47405, USA
[b]Department of Finance, Kelley School of Business, Indiana University, Bloomington, IN 47405, USA
[c]Department of Finance, National Chung Cheng University, Ming-Hsiung, Chia-Yi, 621 Taiwan

Abstract

Despite abundant research which focuses on estimating the level of return on stock market index, there is a lack of studies examining the predictability of the direction/sign of stock index movement. Given the notion that a prediction with little forecast error does not necessarily translate into capital gain, we evaluate the efficacy of several multivariate classification techniques relative to a group of level estimation approaches. Specifically, we conduct time series comparisons between the two types of models on the basis of forecast performance and investment return. The tested classification models, which predict direction based on probability, include linear discriminant analysis, logit, probit, and probabilistic neural network. On the other hand, the level estimation counterparts, which forecast the level, are exponential smoothing, multivariate transfer function, vector autoregression with Kalman filter, and multilayered feedforward neural network. Our comparative study also measures the relative strength of these models with respect to the trading profit generated by their forecasts. To facilitate more effective trading, we develop a set of threshold trading rules driven by the probabilities estimated by the classification models. Empirical experimentation suggests that the classification models outperform the level estimation models in terms of predicting the direction of the stock market movement and maximizing returns from investment trading. Further, investment returns are enhanced by the adoption of the threshold trading rules.

Keywords: Forecasting; Multivariate classification; Stock index; Trading strategy

1. Introduction

Trading in stock market indices has gained unprecedented popularity in major financial markets around the world. The increasing diversity of financial index related instruments, along with the economic growth enjoyed in the last few years, has broadened the dimension of global investment opportunity to both individual and institutional investors. There are two basic reasons for the success of these index trading vehicles. First, they provide an effective means for investors to hedge against potential market risks. Second, they create new profit making

*Corresponding author. Tel.: +1-812-855-8440; fax: +1-812-855-8679.

E-mail addresses: mleung@indiana.edu (M.T. Leung), hdaouk@indiana.edu (H. Daouk), finasc@ccunix.ccu.edu.tw (A.-S. Chen)

0169-2070/00/$ – see front matter
PII: S0169-2070(99)00048-5

opportunities for market speculators and arbitrageurs. Therefore, being able to accurately forecast stock market index has profound implications and significance to researchers and practitioners alike.

Although there exists a vast number of articles addressing the predictability of stock market return as well as the pricing of stock index financial instruments (e.g. S&P 500 futures), most of the proposed models rely on accurate forecasting of the level (i.e. value) of the underlying stock index or its return. In most cases, the degree of accuracy and the acceptability of certain forecasts are measured by the estimates' deviations from the observed values.

Depending on the trading strategies adopted by investors, forecasting methods based on minimizing forecast error may not be adequate to meet their objectives. In other words, trading driven by a certain forecast with a small forecast error may not be as profitable as trading guided by an accurate prediction of the direction of movement (or sign of return.) Therefore, predicting the direction of change of the stock market index and its return is also significant in the development of effective market trading strategies.

In recent years, there has been a growing number of studies looking at the direction or trend of movements of various kinds of financial instruments (such as Maberly, 1986; Wu & Zhang, 1997; O'Connor, Remus & Griggs, 1997). However, none of these studies provide a comparative evaluation of different classification techniques regarding their ability to predict the sign of the index return. Given this notion, we examine various forecasting models based on multivariate classification techniques and compare them with a number of parametric and nonparametric models which forecast the level of the return. Specifically, the major contributions of this study are: (1) to demonstrate and verify the predictability of stock index direction using classification models; (2) to compare the performance of various multivariate classification techniques relative to some econometric and artificial intelligence forecasting techniques; and (3) to develop effective trading strategies guided by the directional forecasts and to test the relative performance of these investment schemes.

The remainder of this paper is organized as follows: A literature review and the background of this study are summarized in the next section. In Section 3, we provide the conceptual foundation of the tested multivariate classification models: parametric linear discriminant analysis, logit, probit, and the nonparametric probabilistic neural network. In addition, we review several level-based forecasting models (adaptive exponential smoothing, Bayesian vector autoregression, multivariate transfer function, and multilayered feedforward neural network) which will be tested against their classification counterparts. Then, we discuss the design of the experiment in Section 4. This section also outlines the proposed index trading strategies and how to apply the directional forecasts and the posterior probabilities supplied by the classification models.

2. Background

2.1. Predictability of stock market returns

'Stock prices do not follow random walks' is the title of a heavily cited paper by Lo and MacKinlay (1988). These authors claim that considerable evidence exists and show that stock returns are to some extent predictable. Most of the research is conducted using data from well-established stock markets such as the US, Western Europe, and Japan.

For the US, several studies have examined the cross-sectional relationship between stock returns and fundamental variables. Variables such as earnings yield, cash flow yield, book to market ratio, and size have been found to have some power in predicting stock returns in these cross-sectional studies. Basu (1977), Fama and French (1992), and Lakonishok, Shleifer and Vishny (1994) are examples of such cross-sectional studies. These studies in general find positive relationships between stock returns and earnings yield, and between cash flow yield and book-to-market ratio, and a negative relationship between stock returns and size.

For the Japanese stock market, Jaffe and Westerfield (1985) and Kato, Ziemba and Schwartz (1990) find some evidence of predictability in the behavior of daily and intraday patterns in index returns. Kato and Schallheim (1985) document size and seasonal anomalies. Chan, Hamao and Lakonishok (1991) relate cross-sectional differences in returns on Japanese stocks to the underlying behavior of earnings yield, size, book to market ratio, and cash flow yield.

Fundamental variables are not the only type of cross-sectional variables that contain information for predictability. DeBondt and Thaler (1985, 1987), Chopra, Lakonishok and Ritter (1992) document that a stock's ranking in terms of its performance relative to the market can contain predictability. Extreme losers have been shown to outperform the market over subsequent years.

In time-series analysis, Fama and French (1993) identify three common risk factors, the overall market factor, factors related to firm size and book-to-market equity which seem to explain average returns on stocks and bonds. Fama and Schwert (1977), Campbell (1987), and Fama and French (1988a,b) find macroeconomic variables such as short-term interest rates, expected inflation, dividend yields, yield spreads between long and short-term government bonds, yield spreads between low grade bonds and high grade bonds, lagged stock price-earnings ratios, and lagged returns have some power to predict stock returns. Chen, Roll and Ross (1986) find that changes in aggregate production, inflation, the short-term interest rates, the slope of term structure as measured by the return difference between long-term and short-term government bonds, and the risk premium as measured by the return difference between low grade bonds and high grade bonds are other macroeconomic factors that have some power to predict stock returns. Also, predictability is not limited to stock returns. For example, Campbell and Mankiw (1987) and Cochrane (1988) find predictability in the GNP, and Huizinga (1987) finds predictability in exchange rates.

Furthermore, Chen et al. (1986) suggest that expected returns are a function of business conditions in that the expected stock market premium is negatively related to the recent growth of economic activity proxying for the health of the current economy and positively related to the expected future growth of economic activity and its conditional variance. Chen (1991) studies the relation between changes in financial investment opportunities and changes in the economy. He provides additional evidence that variables such as the default spread, the term spread, the 1-month T-bill rate, the lagged industrial production growth rate, and the dividend–price ratio are important determinants of future stock market returns. He interprets the ability of these variables to forecast future stock market returns in terms of their correlations with changes in the macroeconomic environment.

The empirical evidence reviewed above is an illustration of large body of theoretical work that deals with Arbitrage Pricing Models. Ross (1976) introduces the Arbitrage Pricing Theory (APT) and derives a model where stock returns are a linear function of a set of state variables (factors). The APT does not imply predictability per se, since the factors are assumed to be observed at the same time returns are. A stream of literature called Conditional Asset Pricing

Models has worked on deriving models that uses some available information set to explain assets returns. Ferson and Harvey (1991) show that predictability in stocks returns are not necessarily due to market inefficiency or overreaction from irrational investors, but rather to predictability in some aggregate variables that are part of the information set. More specifically, they argue that stock returns are predictable because macro variables (such as interest rates or consumption growth) which determine to a certain extent stock returns, are themselves predictable.

It is a well-established practice in the recent empirical finance literature to account for the predictability of stock returns given the investors information set by using macroeconomic variables that are public information. For example, an event study of insider trading by Ferson and Schadt (1996) shows that the omission of variables like lagged stock returns and previous interest rates could lead to misleading results. Stock returns predictability given aggregate variables in the investors information set is a well-accepted fact. The question that remains is how to use the information set in an optimal way for forecasting and trading.

2.2. Forecasting the direction of index return

Most trading practices adopted by financial analysts rely on accurate prediction of the price levels of financial instruments. However, some recent studies have suggested that trading strategies guided by forecasts on the direction of the change in price level are more effective and may generate higher profits. Wu and Zhang (1997) investigate the predictability of the direction of change in the future spot exchange rate. In another study, Aggarwal and Demaskey (1997) find that the performance of cross-hedging improves significantly if the direction of changes in exchange rates can be predicted. Based on the S&P 500 futures, Maberly (1986) explores the relationship between the direction of interday and intraday price change. O'Connor et al. (1997) conduct a laboratory-based experiment and conclude that individuals show different tendencies and behaviors for upward and downward series. This further demonstrates the usefulness of forecasting the direction of change in the price level, that is, the importance of being able to classify the future return as a gain or a loss. The findings in these studies are reasonable because an accurate point estimation, as judged by its deviation from the actual observation, may not be a good predictor of the direction of change in the instrument's price level. This is a common case for the traders in that the predicted direction will immediately affect their decisions on buying or selling the instrument. Finally, in their study on the All Ordinaries Index futures traded at the Australian Associated Stock Exchanges, Hodgson and Nicholls (1991) suggest conducting an evaluation of the economic significance of the direction of price changes in future research.

2.3. Model inputs

Studies related to the predictability of return and its determinant factors are ample and one can easily find such studies in the literature. Hence, for the sake of brevity, we will omit the discussion of economic rationale in this paper. Table 1 displays the set of potential macroeconomic input variables which are used by the forecasting models analyzed in this paper. This study confines the potential input variables to interest rates, consumer price index, industrial production, and lagged returns. These are the most easily available input variables that are observable to a forecaster. Though other macroeconomic variables can be used as inputs, the general consensus in the literature is that the majority of useful information for forecasting is subsumed by the interest rates and lagged returns.

Table 1
List of potential macroeconomic input variables and forecasted output variables

Input variables
ST Short Term Interest Rate First difference of 3-month T-bill rate for the US, and first difference of call money rate for the UK and Japan.
LT Long Term Interest Rate First difference of long term government bond rate for the US, first difference of 20-year government bond rate for the UK, and first difference of long term government bond rate for Japan.
R Lagged Index Returns Lagged terms of the continuously compounded excess returns of the broad market index for the different countries, respectively.
CPI Consumer Price Level First difference of consumer price index for the three countries, respectively.
IP Industrial Production Level First difference of industrial production for the three countries, respectively.
Output variables
R Returns on Index Continuously compounded excess returns of the stock market index (S&P 500, FTSE 100, and Nikkei 225) estimated by level estimation models.
C Probability Given the Direction of Return Probabilities estimated by the classification models that the excess return will be positive in the next period.

3. Forecasting index returns

3.1. Data

The financial and macroeconomic data set used in this study is obtained from the TSM data base and the Citibase compiled by DSC Data Services, Inc. and Citicorp Economic Database, respectively. The entire data set covers the period from January 1967 to December 1995, a total of 348 months of observations. The data set is divided into two periods: the first period runs from January 1967 to December 1990 (288 months of observations) while the second period is from January 1991 to December 1995 (60 months of observations.) The first period, which is assigned to in-sample estimation, is used to determine the specifications of the models and parameters for the forecasting techniques. It also serves the purpose of validating the estimated models. The second period is reserved for out-of-sample evaluation and comparison of performances between various forecasting models.

To test the robustness of the classification models, three globally traded broad market indices — S&P 500 for the US, FTSE 100 for the UK, and Nikkei 225 for Japan, are examined. The forecasted variables are the continuously compounded 1 month excess returns of these indices. As it is shown in Eq. (1), the excess return on an index is defined as the continuously compounded return on the price index minus the riskfree interest rate:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) - r_{t-1} \tag{1}$$

where P_t is the price of the stock index traded at period t and r_t is the US riskfree (1-month T-bill) interest rate in period t. Dividends are ignored for this study. The reason to forecast the excess returns (rather than the index levels or the total returns) is that they provide a measure of how well our models perform relative to the minimum returns gained from depositing the money in a riskfree account. In addition, our study assumes that we are in the position of an American investor. This leads to our adoption of US riskfree T bill rate, as opposed to the interest rates of other riskfree instruments issued by foreign governments, in deriving the excess returns. This is because an American investor always has a choice of not investing but simply putting the money into a US account which pays the short term riskfree interest rate.

The independent variables for predicting the index returns are all observable on or before the last day of the month preceding the month to be forecasted. For instance, for the prediction of the index return for March 1990, all independent variables must be observable on or before the last day of February 1990. Constructing the data in this manner ensures that the estimation of out-of-sample forecasts will be similar to the practice in the real world. That is, only observable, but not future, data are used as inputs to the forecasting models.

3.2. Forecasting by classification models

In this section, we briefly summarize the classification models used in this comparative study. Although these models are based on different statistical techniques, they share a common trait — the ability to generate the probability of group membership. In other words, these models are able to estimate the probability of an upward (or downward) movement in the stock index and thus provide a recommendation for trading.

3.2.1. Linear discriminant analysis

Discriminant analysis is a multivariate statistical technique that investigates the differences between two or more groups of observations with respect to a set of independent (input) variables. These independent variables, called discriminator variables, are used to distinguish the characteristics among different groups. In the discriminant analysis, these discriminator variables are combined to form a set of mathematical equations, known as the classification functions. There is a classification function for each group of observations. The classification function which yields the highest Z score indicates the group membership of the input vector to be classified. At the same time, a probability based on the Z scores can also be calculated to determine the most likely group membership.

In this study, we conduct Fisher's discriminant analysis[1] to classify the direction of excess return. Interested readers should refer to Hair, Anderson, Tatham and Black (1995) or other statistical texts for a more detailed description. After the process of model selection and validation, the Fisher discriminant analysis yields the following classification functions based on the in-sample data from January 1967 to December 1990 (the notations follow the ones explained in Table 1). The forecast is classified as an upward movement (a positive return) if $Z_{pos} > Z_{neg}$, that is, the Z score of the classification function for

[1] Linear discriminant analysis in our experiment is performed by the SPSS statistical package.

positive return is larger than the one for negative return:

US S&P 500

$$Z_{\text{pos}} = 5.334R^{\text{SP}}_{t-1} - 0.268ST^{\text{US}}_{t-1} - 0.730 \tag{2}$$

$$Z_{\text{neg}} = -7.080R^{\text{SP}}_{t-1} + 0.301ST^{\text{US}}_{t-1} - 0.749 \tag{3}$$

UK FTSE 100

$$\begin{aligned} Z_{\text{pos}} = {} & 5.926R^{\text{FTSE}}_{t-1} - 1.875R^{\text{FTSE}}_{t-2} \\ & + 1.925R^{\text{FTSE}}_{t-3} + 1.769R^{\text{FTSE}}_{t-4} \\ & + 0.129LT^{\text{UK}}_{t-1} + 2.262CPI^{\text{UK}}_{t-1} - 1.070 \end{aligned} \tag{4}$$

$$\begin{aligned} Z_{\text{neg}} = {} & -3.306R^{\text{FTSE}}_{t-1} + 2.274R^{\text{FTSE}}_{t-2} \\ & - 0.656R^{\text{FTSE}}_{t-3} - 1.009R^{\text{FTSE}}_{t-4} \\ & + 0.515LT^{\text{UK}}_{t-1} + 2.675CPI^{\text{UK}}_{t-1} - 1.196 \end{aligned} \tag{5}$$

Japan Nikkei 225

$$\begin{aligned} Z_{\text{pos}} = {} & 9.107R^{\text{NIK}}_{t-1} - 1.194R^{\text{NIK}}_{t-2} + 1.914R^{\text{NIK}}_{t-3} \\ & + 3.490R^{\text{NIK}}_{t-4} - 0.763 \end{aligned} \tag{6}$$

$$\begin{aligned} Z_{\text{neg}} = {} & -7.885R^{\text{NIK}}_{t-1} + 4.533R^{\text{NIK}}_{t-2} \\ & + 0.053R^{\text{NIK}}_{t-3} + 0.876R^{\text{NIK}}_{t-4} - 0.733 \end{aligned} \tag{7}$$

These estimated functions are then used to determine the sign of return for each monthly out-of-sample period. In our experiment, the discriminant scores Z are also used to compute $\hat{C}_t$, the posterior probability of group classification for the sign of return in the next period. The mathematical operation can be found in Hair et al. (1995).

3.2.2. Binary choice models: logit and probit

Binary choice models are appropriate to use when trying to model dependent variables that can take on only binary values (e.g. 0 or 1). The general form of the model is:

$$P(Y_i = 1|X_i) = F(X_i\beta) \tag{8}$$

where the dependent variable Y takes the value of either 0 or 1. The question hinges on the value of the parameter P, the probability that Y equals one. X is the set of explanatory variables and $F(\cdot)$ is a nonlinear function of the conditional mean. The only difference between the logit and probit models is the functional form of $F(\cdot)$. $F(\cdot)$ is the cumulative density function (CDF) of the logistic distribution for the logit model, whereas it is the CDF of the normal distribution if a probit model is used. Readers should check out Greene (1993) for a more thorough description of the method.

Binary choice models are particularly suited for our study since we want to predict the direction in the movement of a stock index. The direction of a movement is binary in nature (up or down). Therefore, the logit and probit models will permit us to calculate the probability of an up or down move given all the explanatory variables in our information set[2]. Based on the in-sample data, the chosen logit model specifications are as follows:

US S&P 500

$$\begin{aligned} \hat{C}_t = F(& 0.058 + 12.445R^{\text{SP}}_{t-1} - 6.116R^{\text{SP}}_{t-2} \\ & + 4.686R^{\text{SP}}_{t-3} - 5.130R^{\text{SP}}_{t-4} + 5.390R^{\text{SP}}_{t-5} \\ & - 0.643ST^{\text{US}}_{t-1} + 0.052ST^{\text{US}}_{t-2} - 0.152ST^{\text{US}}_{t-3} \\ & - 0.270ST^{\text{US}}_{t-4} - 0.368ST^{\text{US}}_{t-5}) \end{aligned} \tag{9}$$

UK FTSE 100

$$\begin{aligned} \hat{C}_t = F(& 0.135 + 10.607R^{\text{FTSE}}_{t-1} - 4.794R^{\text{FTSE}}_{t-2} \\ & + 2.183R^{\text{FTSE}}_{t-3} + 2.815R^{\text{FTSE}}_{t-4} - 0.054ST^{\text{UK}}_{t-1} \\ & + 0.028ST^{\text{UK}}_{t-2} + 0.078ST^{\text{UK}}_{t-3} + 0.052ST^{\text{UK}}_{t-4} \\ & + 0.009ST^{\text{UK}}_{t-5} - 0.006ST^{\text{UK}}_{t-6} \\ & + 0.172ST^{\text{UK}}_{t-7}) \end{aligned} \tag{10}$$

[2]Logit and probit estimations are performed by RATS computer package.

Japan Nikkei 225

$$\hat{C}_t = F(0.291 + 16.815R^{\mathrm{NIK}}_{t-1} - 4.773R^{\mathrm{NIK}}_{t-2} + 2.241R^{\mathrm{NIK}}_{t-3} - 0.649LT^{\mathrm{JAP}}_{t-1} - 0.035ST^{\mathrm{JAP}}_{t-2} - 1.057LT^{\mathrm{JAP}}_{t-3}) \quad (11)$$

The estimated probit models are:

US S&P 500

$$\hat{C}_t = G(0.039 + 7.557R^{\mathrm{SP}}_{t-1} - 0.879R^{\mathrm{SP}}_{t-2} + 3.141R^{\mathrm{SP}}_{t-3} - 2.976R^{\mathrm{SP}}_{t-4} + 4.737R^{\mathrm{SP}}_{t-5} - 0.557ST^{\mathrm{US}}_{t-1} + 0.023ST^{\mathrm{US}}_{t-2} - 0.085ST^{\mathrm{US}}_{t-3} - 0.210ST^{\mathrm{US}}_{t-4} - 0.205ST^{\mathrm{US}}_{t-5}) \quad (12)$$

UK FTSE 100

$$\hat{C}^t = G(0.154 + 8.624R^{\mathrm{FTSE}}_{t-1} - 1.981R^{\mathrm{FTSE}}_{t-2} + 0.451R^{\mathrm{FTSE}}_{t-3} + 4.018R^{\mathrm{FTSE}}_{t-4} - 0.031ST^{\mathrm{UK}}_{t-1} + 0.004ST^{\mathrm{UK}}_{t-2} + 0.067ST^{\mathrm{UK}}_{t-3} + 0.083ST^{\mathrm{UK}}_{t-4} - 0.015ST^{\mathrm{UK}}_{t-5} + 0.016ST^{\mathrm{UK}}_{t-6} + 0.105ST^{\mathrm{UK}}_{t-7}) \quad (13)$$

Japan Nikkei 225

$$\hat{C}_t = G(0.219 + 9.543R^{\mathrm{NIK}}_{t-1} - 5.451R^{\mathrm{NIK}}_{t-2} + 0.287R^{\mathrm{NIK}}_{t-3} - 0.419LT^{\mathrm{JAP}}_{t-1} + 0.077ST^{\mathrm{JAP}}_{t-2} - 0.417LT^{\mathrm{JAP}}_{t-3}) \quad (14)$$

The estimated probability $\hat{C}_t$ that we would realize a positive return in the next period is a logistic CDF $F(\cdot)$ of explanatory variables if a logit model is used in forecasting. Likewise, if a probit model is used instead, $\hat{C}_t$ is the probability derived from a normal CDF.

3.2.3. Probabilistic neural networks

In addition to the classification models described in the last two sections, nonparametric neural network models[3] are also used to predict the sign of index return. Our classification network models[4] are based on the Probabilistic Neural Network (PNN) proposed by Specht (1988, 1990). A few studies such as Yang (1999) and Kim (1998) which employ PNN for financial forecasting have shown some promising results. PNN is conceptually built on the Bayesian method of classification. Theoretically, given enough information, the Bayesian method can classify a new sample with the maximum probability of success (Wasserman, 1993, p. 37). A complete description of PNN and its mathematical background can also be found in the same text.

As explained in Section 3.1, our network training and validation scheme involves dividing the first period data into two segments. The first segment is used to train the network whereas the second segment is used to determine and validate the network architecture and model specification. For the sake of brevity, details of this procedure are not reported here. Enthusiastic readers can obtain the information from the authors or refer to Chen and Leung (1998). The selected models for out-of-sample forecasting/evaluation have the following functional forms:

US S&P 500

$$\hat{C}^{\mathrm{SP}}_t = F(R^{\mathrm{SP}}_{t-1}, ST^{\mathrm{US}}_{t-1}) \quad (15)$$

UK FTSE 100

$$\hat{C}^{\mathrm{FTSE}}_t = G\left(R^{\mathrm{FTSE}}_{t-1}, R^{\mathrm{FTSE}}_{t-2}, R^{\mathrm{FTSE}}_{t-3}, R^{\mathrm{FTSE}}_{t-4}, LT^{\mathrm{UK}}_{t-1}\right) \quad (16)$$

[3]Many neural network applications are related to financial decision making. Hawley, Johnson and Raina (1990) and Refenes (1995) provide an overview of neural network models used in the fields of finance and investment.

[4]Readers who are interested in the program codes should refer to Masters (1995).

Japan Nikkei 225

$$\hat{C}_t^{\mathrm{NIK}} = H\left(R_{t-1}^{\mathrm{NIK}}\right) \quad (17)$$

where $\hat{C}_t$ is the posterior probability of realizing a positive return in the next period, and $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ represent the functional relationships between the dependent and independent variables estimated by the PNN.

3.3. Forecasting by level estimation models

3.3.1. Adaptive exponential smoothing

Makridakis, Wheelwright and McGee (1983) and Mabert (1978) described an extension to traditional exponential smoothing model, generally known as adaptive exponential smoothing. This approach continuously evaluates the performance in the previous period and updates the smoothing coefficient. The form of the adaptive exponential smoothing model[5] is similar to that of the simple single exponential smoothing model:

$$\hat{R}_{t+1} = \alpha_t X_t + (1-\alpha_t)\hat{R}_t \quad (18)$$

where $\hat{R}_t$ is the forecast for period t and X_t is the actual observation made in period t, and

$$\alpha_{t+1} = \left|\frac{E_t}{M_t}\right| \quad (19)$$

$$E_t = \beta e_t + (1-\beta)E_{t-1} \quad (20)$$

$$M_t = \beta|e_t| + (1-\beta)M_{t-1} \quad (21)$$

$$e_t = X_t - \hat{R}_t \quad (22)$$

α and β are parameters between 0 and 1 and $|\cdot|$ denotes absolute values.

Based on the results of experiment, the values of β are set to be 0.75, 0.90, and 0.95 for the S&P 500, FTSE, and Nikkei prediction models, respectively. These values are determined by accessing the performance of the models in the first 228 months within the in-sample period (from January 1967 to December 1985). The estimated parameters are then validated using the remaining 60 months (from January 1986 to December 1990) within the in-sample period. Since we assume that historical performance provides meaningful information to predict the future, these values for β are used in the forecasting in the reserved out-of-sample period.

3.3.2. Vector autoregression with Kalman filter updating

The vector autoregression (VAR) has proven to be a successful technique for forecasting systems of interrelated time-series variables in the macroeconomics literature. Notationally, a VAR model with a lag length of p can be represented as:

$$Z_t = \sum_{s=1}^{p} \phi(s)Z_{t-s} + \varepsilon_t \quad (23)$$

$$E(\varepsilon_t \varepsilon_t') = \Sigma \quad (24)$$

where Z_t is an $(n \times 1)$ vector of variables measured at time period t, $\phi(s)$ is an $(n \times n)$ matrix of the coefficients, p is the lag length of the variables, ε_t is an $(n \times 1)$ vector of random disturbances, and Σ is the variance–covariance matrix. Details of the VAR technique can be found in Hamilton (1994).

The specification of the VAR[6] is determined by accessing the performance of alternative VAR specifications in forecasting the last 60 months (from January 1986 to December 1990) of the in-sample period. Based on our experimental results, the following specifications are selected:

US S&P 500

$$Z = \{R^{\mathrm{SP}}\}, \quad p = 3 \quad (25)$$

UK FTSE 100

$$Z = \{R^{\mathrm{FTSE}}, ST^{\mathrm{UK}}\}, \quad p = 2 \quad (26)$$

[5] We implement the adaptive exponential smoothing models using Excel spreadsheet.

[6] VAR estimations are performed by RATS computer package.

Japan Nikkei 225

$$Z = \{R^{\mathrm{NIK}}\}, \quad p = 2 \tag{27}$$

Kalman filter updating is then used to generate 60 monthly forecasts for the reserved out-of-sample period (from January 1991 to December 1995) using the VAR specification chosen.

3.3.3. Multivariate transfer function

A multivariate transfer function model is essentially an ARIMA model with added exogenous variables. It is believed that the addition of exogenous variables can improve the accuracy of our forecasts if these variables help explaining the stock excess returns. Interested readers should refer to Makridakis et al. (1983) for a detailed explanation of this technique. In our experiment, we follow Box and Jenkins' three-stage method (see Appendix A) aimed at selecting an appropriate model for the purpose of estimating and forecasting a time series. The selected models[7] are written as follows:

US S&P 500

ARMA(1, 1) with short term interest rates (ST)

$$R_t = -0.48R_{t-1} + \varepsilon_t + 0.66\varepsilon_{t-1} + \frac{(-0.017)}{(1 - 0.31L + 0.17L^2 - 0.64L^3)} ST_{t-1} \tag{28}$$

UK FTSE 100

MA(1) with long term interest rates (LT)

$$R_t = \varepsilon_t + 0.48\varepsilon_{t-1} + \frac{(-0.0038 - 0.013L)}{(1 + 0.83L - 0.37L^2 - 0.84L^3)} LT_{t-1} \tag{29}$$

[7]Estimation of multivariate transfer function models is performed by RATS computer package.

Japan Nikkei 225

MA(3) with long term interest rates[8] (LT)

$$R_t = \varepsilon_t + 0.38\varepsilon_{t-1} + 0.057\varepsilon_{t-2} + 0.12\varepsilon_{t-3} + \frac{(-0.0062)}{(1 - 0.40L + 0.41L^2 - 1.02L^3)} LT_{t-1} \tag{30}$$

Similar to the adaptive exponential smoothing models, the transfer function models are estimated using the first 228 observations within in-sample period (from January 1967 to December 1985). After these estimated models are validated by the last 60 months of observations in the in-sample period, they are used to generate out-of-sample forecasts from January 1991 to December 1995. In all cases, Industrial Production (IP) and Consumer Price Index (CPI) did not add predictive value to the model.

3.3.4. Multilayered feedforward neural network

Although the essential operations of neural networks are the same (i.e. the networks accept a set of inputs and, through their processing units, produce a corresponding set of outputs), neural networks can appear in many architectural forms. To provide a comparison with the PNN classifier, we test the performance of the multilayered feedforward neural network (MLFN), commonly known as 'backprop' network. Unlike the PNN which suggests a group classification for a given set of inputs, MLFN provides a point estimate or forecast as the output. Readers can refer to Hassoun (1995) and Zhang, Patuwo and Hu (1998) for an explana-

[8]The denominator expression for the lagged LT term has a root inside the unit circle (0.996). This does not violate the stability conditions for the transfer function. The reason is that there are no AR terms for R. It suffices for R and LT to be stationary for the model to be stable. Results from stationarity tests show that R and LT are indeed stationary (a more complete analysis is available upon request from the authors).

tion of neural network model. Wasserman (1993) also provides a description and comparison of the architectures and mathematical foundation of PNN and MLFN.

Based on the results from our study, we select the network architecture[9] which leads to consistent and reasonable performance in the validation sample period from January 1986 to December 1990. The selected model[10] specifications are as follows (the notations follow those described in Table 1):

US S&P 500

$$\hat{R}_t^{SP} = F(R_{t-1}^{SP}, R_{t-2}^{SP}, R_{t-3}^{SP}, R_{t-4}^{SP}, ST_{t-1}^{US}, LT_{t-1}^{US}) \quad (31)$$

UK FTSE 100

$$\hat{R}_t^{FTSE} = G(R_{t-1}^{FTSE}, R_{t-2}^{FTSE}, R_{t-3}^{FTSE}, R_{t-4}^{FTSE}, ST_{t-1}^{UK}, LT_{t-1}^{UK}) \quad (32)$$

Japan Nikkei 225

$$\hat{R}_t^{NIK} = H(R_{t-1}^{NIK}, R_{t-2}^{NIK}, R_{t-3}^{NIK}, R_{t-4}^{NIK}, LT_{t-1}^{JAP}) \quad (33)$$

where $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ are the arbitrary functions deduced by the neural networks. The trained networks are then applied to the forecasting of the index returns in the out-of-sample period (January 1991 through December 1995).

4. Simulation study and trading strategies

4.1. Empirical evaluation

Each of the forecasting models described in the last section is estimated and validated by the in-sample data. This model estimation and selection process is then followed by an empirical evaluation which is based on the out-of-sample data covering 60 monthly observations from January 1991 to December 1995. At this stage, the relative performance of the models is measured by two primary criteria:

- number of correct forecasts (hits) of the sign of index return; and
- excess returns obtained from index trading.

A comparison of the performance between the groups of classification and level estimation models can thus be carried out. The excess returns are derived from trading strategies which are driven by the forecasts made by the classification and level estimation models. The logic of these trading strategies and their results will be discussed in the next two sections.

The number of correct forecasts of the sign of return for each forecasting model is reported in Table 2. The corresponding hit ratios are also given. The average hit ratio for the group of all four classification models is 61.67% whereas that for the group of all four level estimation models is 56.11%. The pooled standard error is 0.0130 and the z value is 4.28, suggesting the group of classification models perform significantly better than the group of level estimation models. In addition, we also test the null hypothesis of no predictive effectiveness, that is, whether the hit ratio of a group of models is

[9] An imperative issue in designing a network is determining the appropriate number of units in the hidden layer. Unfortunately, there is no consistent answer to this question. Although a large hidden layer may provide greater flexibility in functional mapping, it may also lead to the problem of overfitting and hamper the predictive strength of the network model. Nevertheless, Salchenberger, Cinar and Lash (1992) offered a simple guideline of which the number of hidden units should be about 75% of the number of input units. This point was echoed by Jain and Nag (1995). Based on the results from our study, we select the network architecture which leads to consistent and reasonable performance in the validation sample period from January 1986 to December 1990. The resulting construct of the network contains a hidden layer with ten hidden units for both the US and UK models. For the Japan model, the network contains one hidden layer with six hidden units.

[10] MLFN estimations are performed by ThinkPro computer package.

Table 2
Comparison of the predictive strength of classification and level estimation models[a]

		US S&P 500		UK FTSE 100		Japan Nikkei 225	
		Number	Ratio	Number	Ratio	Number	Ratio
Classification models	Discriminant analysis	34	0.57	36	0.60*	**41**	**0.68***
	Logit	36	0.60*	36	0.60*	38	0.63*
	Probit	36	0.60*	36	0.60*	38	0.63*
	Probabilistic neural network	**38**	**0.63***	**37**	**0.61***	38	0.63*
Level estimation models	Adaptive exponential smoothing	29	0.48	33	0.55	**38**	**0.63***
	Vector autoregression with Kalman filter	32	0.53	32	0.53	35	0.58
	Multivariate transfer function	32	0.53	**34**	**0.56**	35	0.58
	Multilayered feedforward neural network	**38**	**0.63***	30	0.50	36	0.60*

[a] The table reports the number of times a forecasting model correctly predicts the direction of the index return over the 60 out-of-sample forecast periods from January 1991 through December 1995. A ratio marked with an asterisk (*) indicates a 95% significance level based on a one-sided test of H_0: $p=0.50$ against H_a: $p>0.50$. The best classification and level estimation models for each index are in bold.

significantly different from the benchmark of 0.5. The statistical tests indicate that the hit ratios for both groups of classification and level estimation models are significantly different from 0.5, which justify the capacity of these forecasting models in the prediction of index return. Moreover, the results implies that the level estimation models are useful in the forecasting of the sign of return although they do not perform as well as the classification models.

4.2. Simple threshold trading strategy

Before we present our trading strategies, we need to describe the operational details of the trading simulation. The trading simulation assumes that, in the beginning of each monthly period, the investor makes an asset allocation decision of whether to shift assets into T-bills or stock index fund. It should be noted that the price of the stock index fund is directly proportional to the index level. Further, it is assumed that the money that has been invested in either T-bills or stock index fund becomes illiquid and remains 'locked up' in that security until the end of the month. The horizon of the trading simulation runs from January 1991 to December 1995, the same out-of-sample period used in the forecasting comparison. This gives a total of 60 monthly periods. In the beginning of each month, an investor has to decide to purchase the index fund and short the T-bill, or to short the index fund and purchase the T-bill, depending on the forecast of excess return in the next period. Investing in this fashion does not require initial capital, making all trading strategies comparable on an equal basis.

The trading strategies between the classification and level estimation models are slightly different to take into account of the unique nature related to each type of forecasts. For the classification models, let $\hat{C}_{t+1}$ be the estimated posterior probability of positive return (upward movement) for the period $t+1$. On the other hand, let $\hat{R}_{t+1}$ denote the forecast of excess return to be realized at the end of period $t+1$, if the level estimation models are used. Thus, the trading strategies (i.e. decision rule) can be expressed as follows:

Classification models

If ($\hat{C}_{t+1}>0.5$) then
  Purchase index fund and short T-bill
Else if ($\hat{C}_{t+1}\leq 0.5$) then
  Short index fund and purchase T-bill

Level estimation models

If $(\hat{R}_{t+1} > 0)$ then
  Purchase index fund and short T-bill
Else if $(\hat{R}_{t+1} \leq 0)$ then
  Short index fund and purchase T-bill

Using these trading strategies, we can compute the excess return over all 60 out-of-sample periods (from January 1991 through December 1995) for each forecasting model. Table 3 tabulates the returns from trading the S&P 500, FTSE 100, and Nikkei 225 indices. The average returns over each type of forecasting models (classification versus level estimation) are also presented. The results show that the return based on trading guided by classification models is 54.28%, compared to a return of 36.31% for level estimation models. The difference of 17.97% is approximately equal to an annualized 3.59% of excess return. For comparison purpose, we also compute the excess return for a 'pure index' trading strategy, i.e, always buy the index. This investment scheme yields a total of 13.67% over the 60 out-of-sample monthly periods, which is lower than the returns obtained from the two types of forecasting models.

A comparison of the average returns over each type of models also shows that trading strategies relied on classification scheme are more profitable than those driven by level estimates although the relative performance varies from index to index. Among all three indices, Nikkei trading realizes the most benefit from using classification forecasting instead of level estimation with respect to the average return (increases from 58.43 to 82.57%.) However, the largest gain in return (from 35.62 to 66.56%) is found in the trading of FTSE 100 when the best model from each category is compared. These observations seem to indicate that the Japanese and British markets may be less efficient than the US market.

The last column in Table 3 indicates the average of returns across all three indices for each forecasting model. Similar to the hit rates, the returns generated by neural network models are better than the other models within the same groups. Probabilistic Neural Network (PNN), which yields a total excess return of 64.10%, is the best performer among the forecasting models evaluated in this study. This observation suggests some merit of evaluating the PNN at a more vigorous level.

Table 3
Excess return from index trading guided by the classification and level estimation forecasts[a]

		US S&P 500	UK FTSE 100	Japan Nikkei 225	Average for Each Model
Classification models	Discriminant analysis	44.1	28.2	**86.1**	52.8
	Logit	40.7	22.9	84.1	49.2
	Probit	40.7	28.1	84.1	51.0
	Probabilistic neural network	**49.8**	**66.6**	76.0	**64.1**
	Average for four classification models	43.8	36.5	82.6	–
	Grand average		54.3		
Level estimation models	Adaptive exponential smoothing	28.6	30.0	67.8	42.1
	Vector autoregression with Kalman filter	14.2	10.5	49.4	24.7
	Multivariate transfer function	25.8	23.5	47.7	32.3
	Multilayered feedforward neural network	**33.7**	**35.6**	**68.9**	**46.1**
	Average for four level estimation models	25.6	24.9	58.4	–
	Grand average		36.3		

[a] The table reports the excess return (%) from index trading over all 60 out-of-sample monthly periods from January 1991 through December 1995. The best classification and level estimation models for each category are in bold.

4.3. Multiple threshold trading strategies

The trading strategies discussed above involve a single threshold which determines the course of trading. In other words, if the probability forecast $\hat{C}_{t+1}$ made by a classification model is greater than the threshold of 0.5, the likelihood of realizing a positive return is higher and an investor should buy the index fund to capture the appreciation. Otherwise, a negative is more likely and an investor should buy the T-bills to earn the interest. The same logic holds for the trading strategy driven by a level estimate $\hat{R}_{t+1}$ except the threshold to trigger the trading is zero. Since the results show that directional forecasts generally outperforms the level estimates, we attempt to enhance the trading strategy by better utilizing the information (i.e. $\hat{C}_{t+1}$) supplied by the classification models. The posterior classification probabilities, which can be viewed as indicators of forecast reliability, are compared against multiple levels of threshold. Intuitively, the more $\hat{C}_{t+1}$ is larger than 0.5, the more the classification model is confident about a positive return in the next period. Conversely, the model becomes more confident about a negative return when$\hat{C}_{t+1}$ decreases from 0.5. For the multiple threshold trading strategies, the course of trading is not determined by a single threshold at 0.5 but two thresholds away from this midpoint. Let γ, where $0.5 \leq \gamma \leq 1$, be the coefficient of reliability an investor would like to maintain in his trading. It can be viewed as the degree of an investor's confidence on the predictive strength of the forecasting models or the risk level he would like to take. The smaller the value of γ, the less confidence or risk an investor exhibits. Thus, the multiple threshold trading can be summarized as:

If ($\hat{C}_{t+1} > 0.5/\gamma$) then
  Purchase index fund and short T-bill
Else if ($\hat{C}_{t+1} < 1 - (0.5/\gamma)$) then
  Short index fund and purchase T-bill
Else
  Do nothing (no trading)

This trading scheme consists of the two courses of action used in simple threshold trading and the option of no trading. $(0.5/\gamma)$ and $1-(0.5/\gamma)$ are essentially the thresholds to trigger the

Table 4
Difference of excess returns between the simple and multiple threshold trading strategies at various levels of reliability[a]

Coefficient of reliability	Difference (%)	Coefficient of reliability	Difference (%)	Coefficient of reliability	Difference (%)
1.00	0.00	0.88	−0.14	0.76	0.24
0.99	−0.04	0.87	−0.12	0.75	0.71
0.98	−0.05	0.86	−0.11	0.74	0.75
0.97	−0.09	0.85	−0.11	0.73	0.79
0.96	−0.07	0.84	−0.19	0.72	0.54
0.95	−0.05	0.83	−0.03	0.71	0.75
0.94	−0.05	0.82	−0.07	0.70	0.77
0.93	−0.05	0.81	0.02	0.69	0.74
0.92	−0.08	0.80	0.16	0.68	1.03
0.91	−0.09	0.79	0.12	0.67	1.32
0.90	−0.12	0.78	0.12	–	–
0.89	−0.15	0.77	0.20	–	–

[a] Difference of excess returns is computed as the excess return per trade obtained from simple threshold trading subtracting the excess return per trade obtained from multiple threshold trading over the 60 out-of-sample monthly periods. Coefficient of reliability is used by an investor to align the confidence level of forecast with his risk level in multiple threshold trading.

buying and selling of index fund, respectively. Any probability higher or lower than these thresholds suggests a high degree of confidence with respect to the investor's desired reliability levels. If $(0.5/\gamma) \leq \hat{C}_{t+1} \leq 1 - (0.5/\gamma)$, the forecasting model is not certain about its classification and thus an investor can avoid potential loss by staying out of the market, i.e. making no trade.

Using these multiple threshold trading strategies, we generate the excess return for the out-of-sample period. Results in Table 4 show the difference of excess returns between the simple and multiple trading strategies for various coefficients of reliability γ. The differences are expressed in excess return per trade, instead of overall return in the entire out-of-sample period. It is because the number of trades varies with the value of γ and it is not fair to compare across the returns with different number of trades. Fig. 1 plots the difference in excess returns versus the coefficient of reliability γ. The choice of γ by the trader is an empirical issue. It is reasonable to think that the optimal range of γ should not be in extreme but rather lies inside the domain of definition (i.e. not on the boundary). If γ is very close to 1 then the multiple threshold trading rule is essentially identical to the simple threshold strategy. This will neutralize any advantage of the multiple threshold rule over the simple threshold rule. In other words, the capacity of multiple threshold rule to screen out unreliable forecasts is eliminated. If γ is very close to 0.5 then the probability forecast $\hat{C}_{t+1}$ needs to be extremely high or low (close to 0 or 1) in order to trigger a

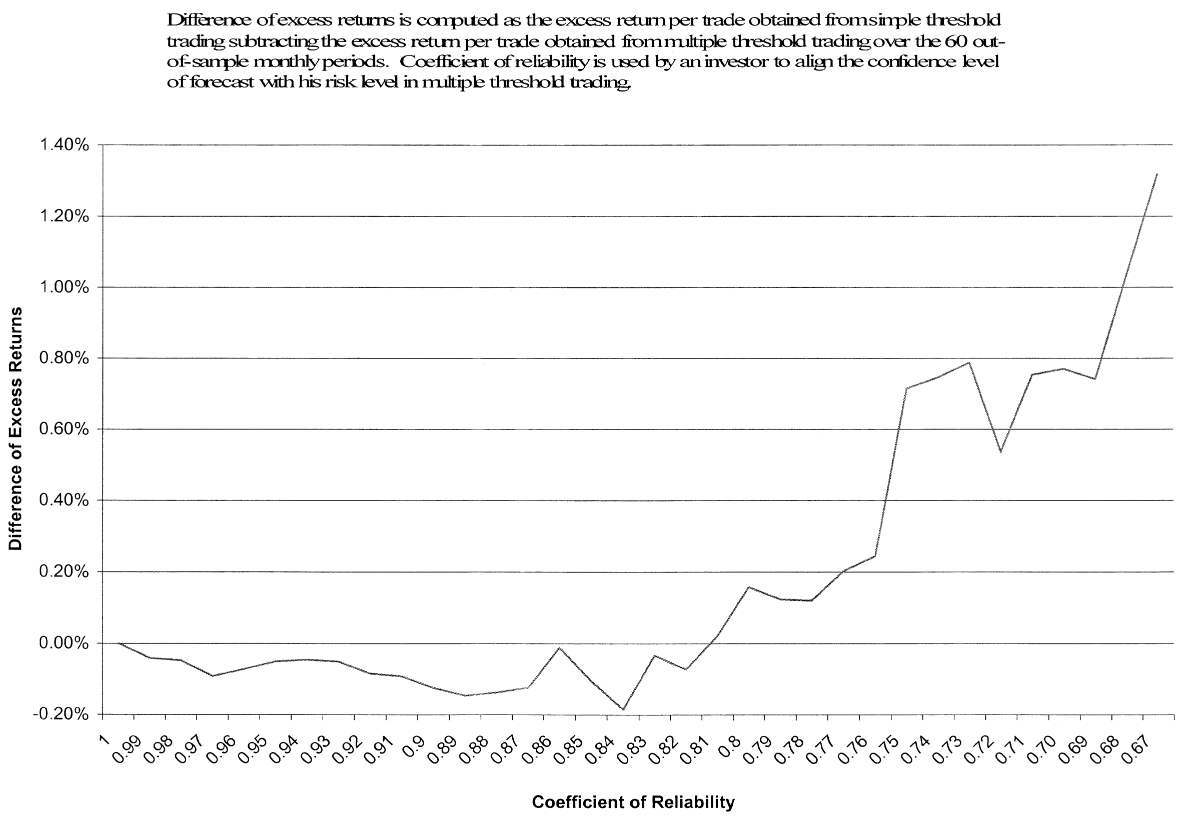

Fig. 1. Difference of excess returns between the simple and multiple threshold trading strategies versus coefficient of reliability.

trade transaction. The consequence is that there will seldom be a trade and therefore not much profit. We hypothesize that the multiple threshold rule will provide the trader with an advantage over the single threshold rule for intermediate values of γ. Fig. 1 offers support to this conjecture. For high values of γ (above 0.8), the difference in excess returns is very close to 0 (actually slightly negative). For this range of γ, the thresholds are too close to 1 to provide the multiple threshold trading rule an advantage over the single threshold counterpart. On the other hand, for low values of γ (less than 0.66), the trading threshold is too close to 0.5 which lead to a total absence of trading in many cases (this explains why the graph is cut off at γ of 0.67). For intermediate values of γ (between 0.67 and 0.75), the multiple threshold rule provides substantially better returns relative to the single threshold rule. In this range, the difference in excess returns is around 0.7% per trade. This return is on a per-trade basis, and thus it represents an excess return of 8.4% on an annualized basis. It is because our simulation assumes that a trade can take place only at the beginning of a month and a trader can make up to 12 trades per year.

The range over which γ should be chosen is, as stated earlier, an empirical issue. Based on the data applied to this study, it is recommended a trader to choose a γ at around 0.7. This corresponds to the upper and lower thresholds of 0.7 and 0.3, respectively. Also, we expect the recommended range of value to hold in general for different assets. A probability above 0.7 (below 0.3) implies that the forecasting method is fairly confident that the index will increase(decrease). Hence, restricting trading to these cases can reduce the number of costly mistakes.

5. Conclusions

We show that the forecasting performance of a group of classification models is superior to that of a group of level estimation models. The classification models included in the study are aimed at forecasting the sign (direction) of index return whereas the level estimation models take the conventional approach to estimate the value of the return. The classification models perform better than their level estimation counterparts in terms of hit rate (number of times the predicted direction is correct). More interestingly, the classification models are able to generate higher trading profits than the level estimation models. This is a clear message for financial forecasters and traders. The message is that their focus should be on accurately predicting the direction of movement as opposed to minimizing the estimates' deviations from the actual observed values. Practitioners should seriously consider incorporating classification models into their forecasting kit. In practice, traders intrigued by our results could use historical returns from the asset they specialize in and test if the use of classification models could have generated higher profits than what they actually obtained from level estimation models. If this is the case, then those traders should at least consider using the classification models in addition to the more traditional models they are using.

In addition, we are able to show that the posterior classification probabilities computed by the classification models can be successfully used in developing 'multiple threshold' trading strategies that enhances trading profits. Under this trading scheme, a trade does not take place unless the forecast is acceptable based on the investor's reliability standard. Also, the empirical test suggests the range of thresholds for which the strategy will work best.

Acknowledgements

The authors would like to acknowledge the helpful comments of the two anonymous referees.

Appendix A

In our experiment, we follow Box and Jenkins' three-stage method aimed at selecting an appropriate model for the purpose of estimating and forecasting a time series:

A.1. Identification stage

We use the SARIMA procedure in RATS statistical software to determine plausible models. The SARIMA procedure uses standard diagnostics such as autocorrelation function (ACF), partial autocorrelation function (PACF), and plots of series.

A.2. Estimation stage

Each of the tentative models is fit and the various coefficients are examined. In this stage, the estimated models are compared using standard criteria such as AIC, SBC, and significance of coefficients.

A.3. Diagnostic checking stage

SARIMA procedure is used to check if the residuals from the different models are white noise. The procedure uses diagnostics tests such as ACF, PACF, Ljung-Box Q-statistic for serial correlation, and Jarque-Bera normality test.

References

Aggarwal, R., & Demaskey, A. (1997). Using derivatives in major currencies for cross-hedging currency risks in Asian emerging markets. *Journal of Futures Markets 17*, 781–796.

Basu, S. (1977). The investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance 32*, 663–682.

Campbell, J. (1987). Stock returns and the term structure. *Journal of Financial Economics 18*, 373–399.

Campbell, J. Y., & Mankiw, N. G. (1987). Permanent and transitory components in macroeconomic fluctuations. *American Economic Review 77*, 111–117.

Chan, L., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan. *Journal of Finance 46*, 1739–1764.

Chen, A., & Leung, M. (1998). Dynamic foreign currency trading guided by adaptive forecasting. *Review of Pacific Basin Financial Markets and Policies 1*, 383–418.

Chen, N. (1991). Financial investment opportunities and the macroeconomy. *Journal of Finance 46*, 529–554.

Chen, N., Roll, R., & Ross, S. (1986). Economic forces and the stock market. *Journal of Business 59*, 383–403.

Chopra, N., Lakonishok, J., & Ritter, J. R. (1992). Measuring abnormal performance: Do stocks overreact? *Journal of Financial Economics 31*, 235–268.

Cochrane, J. H. (1988). How big is the random walk in GNP? *Journal of Political Economy 9*, 893–920.

DeBondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance 40*, 793–805.

DeBondt, W. F., & Thaler, R. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance 42*, 557–581.

Fama, E., & French, K. (1988a). Permanent and temporary components of stock prices. *Journal of Political Economy 96*, 246–273.

Fama, E., & French, K. (1988b). Dividend yields and expected stock returns. *Journal of Financial Economics 22*, 3–25.

Fama, E., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance 47*, 427–465.

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E., & Schwert, W. (1977). Asset returns and inflation. *Journal of Financial Economics 5*, 115–146.

Ferson, W., & Harvey, C. (1991). The variation of economic risk premiums. *Journal of Political Economy 99*, 385–415.

Ferson, W., & Schadt, R. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance 51*, 425–461.

Greene, W. H. (1993). Econometric analysis, 2nd ed, Macmillan, New York.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). Multivariate data analysis, 4th ed, Prentice Hall, Englewood Cliffs, NJ.

Hamilton, J. (1994). Time series analysis, 2nd ed, Princeton University Press, New Jersey.

Hassoun, M. H. (1995). Fundamentals of artificial neural networks, MIT Press, Cambridge, MA.

Hawley, D., Johnson, J., & Raina, D. (1990). Artificial neural systems: A new tool for financial decision-making. *Financial Analysts Journal 23*, 63–72.

Hodgson, A., & Nicholls, D. (1991). The impact of index futures markets on Australian share market volatility. *Journal of Business Finance and Accounting 18,* 267–280.

Huizinga, J. (1987). An empirical investigation of the long run behavior of real exchange rates. *Carnegie Rochester Conference Series on Public Policy 27,* 149–214.

Jaffe, J., & Westerfield, R. (1985). Patterns in Japanese common stock returns: Day of the week and turn of the year effects. *Journal of Financial and Quantitative Analysis 20,* 261–272.

Kato, K., & Schallheim, J. (1985). Seasonal and size anomalies in the Japanese stock market. *Journal of Financial and Quantitative Analysis 20,* 243–260.

Kato, K., Ziemba, W., & Schwartz, S. (1990). Day of the week effects in Japanese stocks. In: Elton, E., & Grubber, M. (Eds.), Japanese capital markets, Harper and Row, New York.

Kim, S. (1998). Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index. *International Journal of Forecasting 14,* 323–337.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance 49,* 1541–1578.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies 1,* 41–66.

Maberly, E. D. (1986). The informational content of the interday price change with respect to stock index futures. *Journal of Futures Markets 6,* 385–395.

Mabert, V. A. (1978). Forecast modification based upon residual analysis: A case study of check volume estimation. *Decision Sciences 9,* 285–296.

Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1983). Forecasting: methods and applications, 2nd ed, Wiley, New York.

O'Connor, M., Remus, W., & Griggs, K. (1997). Going up-going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting 16,* 165–176.

Refenes, A. P. (1995). Neural networks in the capital markets, Wiley, New York.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13,* 341–360.

Specht, D. (1988). Probabilistic neural networks for classification, mapping, or associative memory. In: IEEE International Conference on Neural Networks.

Specht, D. (1990). Probabilistic neural networks. *Neural Networks 3,* 109–118.

Wasserman, P. (1993). Advanced methods in neural computing, Van Nostrand Reinhold, New York.

Wu, Y., & Zhang, H. (1997). Forward premiums as unbiased predictors of future currency depreciation: A non-parametric analysis. *Journal of International Money and Finance 16,* 609–623.

Yang, Z. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research 44,* 67–74.

Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting 14,* 35–62.

Biographies: Mark T. LEUNG is Assistant Professor of Management Science at the University of Texas. He received his BSc and MBA degrees from the University of California, and Master of Business and PhD in operations management from Indiana University. His research interests are in financial forecasting and modeling, applications of AI techniques, scheduling and optimization, and planning and control of production systems. He has published in *Decision Sciences, International Review of Financial Analysis*, *Review of Pacific Basin Financial Markets and Policies, Computers and Operations Research*, and a variety of conference proceedings.

Hazem DAOUK is a PhD candidate in Finance, Kelley School of Business, Indiana University. He received a DESCF from ICS Paris, France, and an MBA from the University of Maryland. His research interests include volatility of asset prices, international finance, and financial econometrics. He has published in the *Journal of Financial Economics and Computers and Operations Research*.

An-Sing CHEN is Professor of Finance at National Chung Cheng University, Taiwan. Dr Chen received his BSc in business from Kent State University, MBA in finance and PhD in Business Economics from Indiana University. His areas of interest are forecasting and modeling, option and derivatives, international finance, and applied econometrics. He has published articles in a variety of journals, including *International Journal of Finance*, *Journal of Investing*, *International Review of Financial Analysis*, *International Review of Economics and Finance, Review of Pacific Basin Financial Markets and Policies and Computers and Operations Research*.